DOCUMENT RESUME

ED 174 671 TM 009 628

AUTHOR Terwilliger, James S.
TITLE Stevens' Scales of Measurement--A Critique and an Alternative.
PUB DATE Apr 79
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Behavioral Science Research; Classification; *Conceptual Schemes; *Measurement; *Measurement Goals; *Measurement Techniques; Psychometrics; Statistical Analysis; Testing; *Typology
IDENTIFIERS *Scales (Measurement); *Stevens (S S)

ABSTRACT

It is asserted that Steven's viewpoint on measurement--that scales are nominal, ordinal, interval, or ratio--has a limited usefulness for behavioral scientists and researchers, and that it is necessary to develop an alternative. Some limitations of Steven's theory are discussed: (1) its assumptions about the nature and purpose of measurement; (2) its limited utility as a guide for the design and interpretation of measurement procedures; and (3) the tendency of Steven's scale types to support fundamental misconceptions about the nature of science. An alternative viewpoint is recommended, and a revised definition of measurement is proposed, whereby the purpose is to index a quantity or amount of an attribute such as interest or aptitude. Rather than using Stevens' four scales of measurement, three ways to index phenomena are suggested: classificaiton systems; ranking techniques; and measurement procedures. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Session 27.14

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED174671

Stevens' Scales of Measurement - A Critique and An Alternative

James S. Terwilliger

University of Minnesota

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

James S. Terwilliger

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting

of the American Educational Research Association,

San Francisco, CA, April, 1979

TM009 628

Introduction

The great majority of current texts on measurement and statistics employed in the behavioral sciences include an early chapter devoted to a discussion of basic definitions of measurement and typologies for classifying variables. The uniformity with which these topics are treated by textbook authors is striking. An observer unfamiliar with the behavioral sciences would quickly conclude that there is a standard "party line" which is the accepted dogma with regard to conceptions of the fundamental nature of measurement. The prevailing viewpoint is, of course, the system first articulated by Stevens (1946, 1951). Steven's definition of measruement, his system for classifying scales of measurement (nominal, ordinal, interval and ratio) and his notions concerning permissible statistical treatments have been dutifully studied by budding behavioral scientists for the past quarter century.

To be sure, serious questions have been raised concerning the viability of Steven's view, especially with regard to the "permissible statistics" issue. There has been a lengthy and often lively debate between those who support Steven's strictures on statistical treatment of data and those who argue against them. Although that debate is interesting it is of secondary interest in the present context. The reader who wishes to pursue the literature on "permissible statistics" is referred to the reveiw by Gardner (1975).

The thesis of this paper is that Steven's viewpoint on measurement has limited utility for both researchers and practitioners in the behavioral sciences. The assumptions implicit in Steven's conception of measurement are, in fact, irrelevant to the primary concerns of most researchers and practitioners. Consequently, there is a need to develop an alternative viewpoint more in line with current theory and practice. The elements of such an alternative have previously been described by Adams (1966) and Rozeboom (1966).

Origins and Assumptions of Steven's Viewpoint

Steven's conception of measurement is a direct extension of Campbell's (1920, 1928) notions. Campbell, a physicist, asserted that the purpose of measurement is to assign numeric values to objects or phenomena in such a way that certain relations among and operations on the assigned values correspond or represent observable relations among and operations on the objects or phenomena to which they are assigned. He recognized that the degree to which a correspondence between numeric values and objects could be achieved depends upon the property of the object being measured. Therefore, he distinguished between extensive properties (quantities) and intensive properties (qualities). The critical difference is that for extensive properties there exists some empirical operation (concantenation) analogous to the operation of addition of numbers while for intensive properties no such operation exists. Weight is an extensive property because when two objects are placed on a scale together their combined weight is equal to the sum of their individual weights. For Campbell, only extensive properties are amenable to "fundamental measurement", i.e., only extensive properties can be measured via direct operations without recourse to prior measurement procedures.

For present purposes the most important feature of Campbell's conception is his premise that the purpose of measurement is to represent real world objects or phenomena. If one accepts this premise, it follows that measurement procedures which establish a more perfect isomorphism between the objects measured and the assigned numeric values are to be preferred over those procedures which establish a less perfect isomorphism. From this perspective, extensive properties which are fundamentally measureable are clearly preferrable to intensive properties which are not because the former can be more perfectly represented. Adams (1966) has referred to this as the "representational theory" of measurement.

Stevens (1946, 1951) acopted the representational viewpoint and further elaborated it as a framework for describing measurement in the behavioral sciences.

Specifically, he suggested the following definition: "Measurement is the assignment of numerals to objects or events according to rules." (1951, p. 1). Since there are many possible rules for assigning numerals to objects and any given rule determines numeric assignments only within a specified level of uniqueness, Stevens further differentiated types of measurement procedures. The result was his four widely cited scale types: nominal, ordinal, interval, and ratio.

In the spirit of representational theory, Stevens argued that there are basic empirical operations which correspond to each scale type as follow:

Scale	Operation
Nominal	Determination of equality
Ordinal	" " greater or less
Interval	" " equality of intervals or differences
Ratio	" " equality of ratios

The operations necessary to create a given scale include all those listed down to and including the operation opposite the scale name.

Where Campbell (1928) had distinguished between extensive and intensive properties based upon the existence of an empirical analog to addition, Stevens distinguishes among four scales of measurement based upon empirical operations which are analogous to various relations existing among members. In both instances, the primary objective is to establish a known degree of correspondence between real world objects and abstract symbolic systems (numbers) such that the relations among the objects are reflected in the relations among the numbers.

It can be seen that the correspondence (or isomorphism) between numbers and objects increases from the nominal scale to the ratio scale because the number of empirical operations needed to create each successive scale is greater and includes all of those needed to create scales above it. Therefore, the nominal scale is commonly regarded as the "weakest" level of measurement and the ratio scale is considered the "strongest" level of measurement.

Stevens gives numerous illustrations of how empirical operations may be employed with physical stimuli to establish nominal, ordinal, interval, and ratio scales in the context of psychophysical experiments. The objective of these studies is to establish a metric for some subjective judgmental or perceptual process using standardized laboratory stimuli and experimental methods, e.g. paired comparison, threshold judgments, rank ordering, fractionation techniques.

It should be noted that a physical scaling of the stimuli is frequently employed to determine the level (ordinal, interval or ratio) at which the subjective (psychological) process can be measured. For example, the units (just noticeable differences) on a subjective scale of "heaviness" are judged to be equal or unequal based upon their correspondence to differences on an analogous physical scaling (actual weights) of the stimuli employed. In such instances, some physical property of the stimuli is the basis for classifying the scale type for the corresponding subjective dimension.

The mathematical rationalization of Steven's theory of scale types has been substantially extended through the work of Scott and Suppes (1958), Suppes and Zinnes (1963), and Pfanzagl (1966). Starting from the representational viewpoint, these writers have developed a detailed axiomatization of scale types employing the concept of "relational systems" for mapping properties of real world objects into number systems. The result is a formal set of proofs associated with each of the Stevens' scale types. The emphasis in this work is establishing the conditions under which the structure of relations among empirical objects is isomorphic to the structure of relations pertaining within number systems.

Limitations of Stevens' Scale Types

Stevens' scale types have dominated thinking about measurement in the behavioral sciences for the past three decades. However, a close examination of the way in which measurement procedures are employed by researchres and professionals reveals a

discontinuity between Stevens' conceptions and the day-to-day applications of measurement to the solution of problems of scientific interest. Many of the weaknesses of representational theory in general and Stevens' scale types in particular have been previously noted by Adams (1966) and Rozeboom (1966). Unfortunately, there is little evidence that these criticism and their implications have been widely understood and appreciated. Therefore, a recapitulation and elaboration of the major concerns is presented here.

a. Stevens' assumptions.

The most fundamental issues are the assumptions made regarding the nature and purpose of measurement in the behavioral sciences. Assuming that psychological phenomena and physical phenomena are capable of definition in a like manner, Stevens and others who adopt the representational viewpoint assert that the purpose of measurement is to describe or characterize properties of "real world" objects or events as accurately as possible with symbols which have quantitative interpretations. The more accurate the description (i.e., the more perfect the isomorphism between the objects and the numbers) the "better" the measurement procedure. Thus, Stevens states:

> When this correspondence between the formal model and its empirical counterpart is close and tight, we find ourselves able to discover truths about matters of fact by examining the model itself. Thus we calculate the flight of a bullet or the course of a comet without laying hands on either. (1951, p. 2).

It is clear that an empirical check is always available in the examples cited by Stevens. The calculations made using the formal model provide a prediction which is subject to confirmation via some experimental procedure or empirical observation. In the case of a bullet or a comet, the phenomena are directly observable (perhaps with the aid of special apparatus) and the properties

involved are instances of Campbell's extensive properties. In later illustrations from psychophysical studies, Stevens typically presents settings in which a subjective (psychological) phenomena is mapped against a physical property of the stimuli employed in the experiment. Again, an empirical check of the mapping can be performed directly by determining whether predicted subjective reactions occur for given stimuli with known physical characteristics. In the rare instances where no physical measurement is directly employed, empirical checks are made by confirming predictions of behavior made directly from subjective measures, e.g. to verify that loudness (measured in sones) is a ratio scale, the loudness of a sound heard in two ears has been shown to be twice the loudness heard with one ear.

Of course, most research in the behavioral sciences does not employ measures of physical properties of objects or events nor do researchers typically have at their disposal direct experimental techniques for verifying measurement procedures. Instead, the researcher usually attempts to index a psychological construct with some fairly arbitrary set of operations and scoring procedures, e.g. rating scales, checklists, tests, etc. employed to index a wide variety of intellectual, personality and biographical characteristics. The motivation for such work is the identification of general lawful relationships which can be verified in subsequent studies with comparable samples of subjects.

Where measurement is used by a professional as an aid to decision-making in an applied setting, it is likewise true that interest most often centers upon psychological traits for which there are no known physical analogs. Again, operational definitions for unobservable phenomena are employed to obtain a set of scorable responses which are converted into a quantitative index. The decision is presumably based upon the previously demonstrated empirical relation (validity) between the quantitative index and alternative outcomes which are possible as a result of the decision. Common examples are the clinical use of tests and inventories for diagnosis and the use of various psychometric instruments for

selection, placement and classification of personnel.

The criterion for judging a measurement procedure should be conceived in terms of the theoretical and applied uses intended for the instrument and not in terms of the relationship between observed scores and some arbitrary definition of "ultimate reality." The goal of science is the establishment of lawful relationships which are expressable in a generalizable form and this goal should not be overlooked in discussions of measurement. This is reflected in Adams (1966) quote of Nagel's statement in 1931:

> Consequently, if we inquire why we measure in physics, the answer will be that if we do measure, and measure in certain ways, then it will be possible to establish the equations and theories which are the goal of inquiry. (p. 152)

A parallel statement concerning the use of measurement for applied decision-making can be formulated from the work of Cronbach and Gleser (1965), viz.

> If we do measure, and measure in certain ways, then it will be possible to arrive at decisions which result in greater benefit to the organization (or individual) than alternatives of comparable cost and thus yield greater utility which is the goal of decision-making.

The representational theory of measurement appears to be oblivious to the purposes for which measures are constructed. In his concern for ascertaining how well a formal measurement procedure "represents" empirical phenomena, Stevens has failed to recognize that: a) in most settings no standard empirical operations exist against which observed quantitative results can be checked and b) even if such operations did exist, the proper criterion for judging the quality of a measurement procedure is its demonstrated relationship to other measures and/or its utility in decision-making. Adams (1966) and Rozeboom (1966)

have previously made similar criticisms of Stevens' position.

It should also be noted that other writers who have taken the representational viewpoint have recognized its serious limitations as a general model for psychological measurement. Suppes and Zinnes (1963) use the term "pointer measurement" to describe numerical assignment based upon direct readings taken from a validated instrument, e.g. measurement of weight on a platform scale. They state:

> All too often in the behavioral sciences a direct reading instrument is available (and used) despite the fact that its readings are not justified; the readings do not correspond to any known fundamental or derived numerical assignment. On the surface it would appear that such pseudo-pointer instruments would be useless and their readings meaningless. Their prevalence in psychology (e.g., mental tests, questionnaires, indices), however, suggests that this conclusion may be too strong. --- A mental test score, for example, based on the number of correct answers may be able to predict success in college or in a job. <u>The justification in the use of such instruments would then lie solely in the degree to which they are able to predict significant events</u>, not as with most "normal" fundamental measures, in the homomorphism between an empirical system and a numerical system. (p. 21, emphasis added).

Pfanzagl (1966) uses the phrase "measurement by fiat" to refer to any situation where there is no clear mapping of an empirical system into a numerical system. He states:

> Measurement by fiat occurs if a prescientific concept turns out to be of great importance and no scaling procedure is available. Typical examples for measurement by fiat are so called intelligence tests, scales for social status, etc. The practical importance of scales like these is proved by their probation for predictions. (p. 32, emphasis added).

b. Limited utility of Stevens' scales

A system of classifying measurement techniques should provide distinctions which are generally agreed upon and be useful as a guide for the design and interpretation of measurement procedures. Stevens' system does neither of these. Attempts to classify commonly employed psychometric techniques as nominal, ordinal, interval or ratio have resulted in substantial confusion. Even Stevens (1951) is uncertain whether "standard scores" on achievement tests should be considered as an interval scale since he lists it as an example followed by "(?)". He also notes that some psychometricians employ the normal curve assumption to derive equal scale units but warns that, "This process smacks of a kind of magic - a rope trick for climbing the hierarchy of scales." Most writers vascillate between "ordinal" and "interval" when attempting to classify standardized measures of aptitude, ability, and achievement. There is a similar ambivalence with respect to how to treat measures of attitude, interest, and personality traits.

The problem is, of course, that no definitive empirical checks exist to test the claim that a given procedure should be classified in a particular way. The unambiguous examples of the scale types (which appear in almost every textbook) are:

Scale	Example
Nominal	"Numbering" football players
Ordinal	Order in which horses finish in a race
Interval	Temperature (Fahrenheit and Centigrade)
Ratio	Length or weight

These examples have no bearing upon the concerns of behavioral scientists and provide no help whatsoever in the construction and interpretation of psychological and educational measures.

Rozeboom (1966) has argued very persuasively against the utility of Stevens' scale types by pointing out that such a classification focuses attention on the wrong issues. He states that the Stevens' viewpoint,

> disastrously inverts the proper order of inquiry by seeking to answer questions about scale-property meaningfulness in terms of the scale's type rather than by judging a scale's type in terms of what on it is meaninful. (p. 196).

And later he concludes,

> In short, the alpha and omega of scaling is significant content. If we have it, then we can formally process it in whatever fashion suits our whim. And if we do not, no incantations about scale types and admissible transformations will provide it for us. (p. 197).

c. Misconceptions fostered by Stevens' scale types

A final major problem with Stevens' position is that it tends to foster fundamental misconceptions about the nature of science. The first is that an "ultimate definition" of reality exists in which variables are measured in some "fundamental" sense. Although not explicitly stated, this is strongly suggested

in Stevens' examples which stress establishing an isomorphism between assigned scale values and specified empirical results or other natural phenomena. The result is an artificial dichotomy between the fundamental or "real" variable and the assigned numeric value which is only an approximation to the "real" variable. Such thinking does not do justice to the complexities of educational and psychological measurement and certainly does not reflect current philosophical views which stress operational definitions as a means for indexing hypothetical constructs. Nunnally (1967) states that he,

> opposes the concept of "real" scales and deplores the confusion which this conception has wrought. It is much more appropriate to think of any measurement scale as a convention - an agreement among scientists that a particular scaling of an attribute is a 'good' scaling. - - - With somes types of measures, the nature of a 'good' scaling is so readily agreed on by all that a convention is easily established. Thus it is with length and weight: no one opposes the ways of getting the zero point and establishing the intervals. (p. 30).

A second pitfall of Stevens' scales is the implicit assumption that a given measurement procedure can be uniquely classified as one scale type. This ignores the fact that the same procedure can be employed for many different purposes and, consequently, may have several different interpretations.

As Jones (1971) states:

> Then the scale type is always relative to a particular interpretation and is not a fixed property of that set of numbers that result from a given measuring operation. For example, the number of dollars received as a measure of monthly wage conforms to a ratio scale type. Yet that same

> number of dollars, if interpreted as an index of utility of income, may meet only the criteria of an ordered scale (since utility of dollars may be a nonlinear transformation of number of dollars). (p. 354)

An Alternative Viewpoint

The representational point of view has been the dominant perspective on measurement despite the limitations cited above because there has been no well defined (and well publicized) alternative position. It is easy to see why this is the case. Stevens' theory of scale types comes in a neat and "easy to digest" package which can be nicely summarized in a table giving the names of the four scales, the mathematical group structure for each scale, the permissible statistics to be employed with each scale (this is an optional feature) and typical examples of each scale (usually the same ones - see previous section). No wonder that several generations of students have faithfully mastered this material and adopted the underlying viewpoint as the definitive statement on measurement!

The first step in formulating an alternative to the representational viewpoint is to adopt a revised definition of measurement which is more consistent with the meaning of that term in actual practice. The following is proposed:

> Measurement is the assignment of numbers to objects according to rules in order to index the quantity of some hypothesized (inferred) attribute of the objects.

This definition is a variation on those given by Rozeboom (1966), Nunnally (1967) and Jones (1971). It differs from Stevens' definition in two important ways. First, the purpose of measurement is to index some quantity or amount of a presumed characteristic or attribute. In the words of Rozeboom (1966), "In the tough sense of the word, 'measurement' is assessment of quantity." This implies that whenever measurement occurs, some unit (however arbitrary) is

employed and the numbers which result are interpreted as having quantitative meaning.

Second, the attributes which are indexed by measurement procedures are typically hypothetical in nature, e.g., aptitude, interest, motivation, etc. Such attributes are defined by a specified set of operations and do not exist independently of the operations by which they are defined. There is no ultimate empirical check on the numbers assigned to objects by a given set of rules and there are no "real" or "fundamental" scales for the attributes. Consequently, it is impossible to speak of an isomorphism between the numbers and the objects to which they are assigned.

Under this definition of measurement there is no such thing as a "nominal scale" because crude classification systems do not index quantity. The same is true of "ordinal scales" which derive from ranking techniques that simply apply numbers as an index of the order in which various objects stand in respect to each other on some attribute. Neither classification nor ranking procedures employ a unit of measurement, hence it is inappropriate to refer to them as measurement. As Jones (1971) states:

> This is not to deny the usefulness of ordered scales or even of nominal scales but simply reflects the view that measurement is an assessment of magnitude and that a measure represents an amount. (p. 354).

Rather than four "scales of measurement", it would be more accurate to speak of three different ways to index phenomena: classification systems, ranking techniques, and measurement procedures. Table 1 distinguishes these three in terms of characteristics, purposes and examples of each.

Measurement procedures yield what is frequently referred to as "metric" data whereas classification and ranking result in "nonmetric" data. Nonmetric

systems are frequently employed simply to distinguish or label objects on the basis of rather simple and well-defined differences as can be seen in the examples of classification and ranking systems in Table 1. Measurement systems are rarely employed in this fashion. As is evident from the examples in Table 1, the aim of measurement is typically some theoretical (hypothetical) attribute which is not readily defined in terms of simple observations, conventions or rules. Therefore, measurement systems are usually much more complex than are classification or ranking systems.

- - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - -

The term "index" was chosen in the above definition rather than the term "represent" to emphasize the arbitrary features associated with any measurement procedure. Many different operational definitions can be formulated for the same hypothetical attribute, each one resulting in quantitative results which provide a different "indexing" of the attribute. This follows Adams' (1966) reasoning that:

> at least one of the principal reasons for making measurements is to provide objective <u>indices</u> of phenomena. This leads to a kind of <u>informational</u> account of measurement, and to the view that measures of a quantity are not so much 'true' or 'false' as they are more or less informative about the phenomena which they are supposed to be indices of. (p. 126).

Adams (1966) has described an information theoretic conceptualization of measurement which he believes is more appropriate than the representational viewpoint. His approach builds upon Shannon and Weaver's (1949) statistical formulation of measures of "uncertainty" and "transmitted information".

Although the details of Adams' (1966) development are beyond the scope of the present paper, he suggests a clear alternative to representational theory that allows for empirical tests of competing measurement operations where such tests are feasible. Unfortunately, the examples Adams (1966) considers are for physical measurement (Mohs scale of hardness and weight measurement) and it is not clear how one can apply his information theory approach to psychological indices.

The major theme in the writings of Adams (1966), Jones (1971), Nunnally (1967) and Rozeboom (1966) is a greater emphasis upon the purpose for which instruments are designed and employed in both theoretical and applied settings. This implies need for more concern for pragmatic issues in measurement typically associated with the process of validation. Ultimately, it is the empirical support which can be mustered for a proposed interpretation of scores on a measure that gives those scores scientific and practical significance. Hence, the nature of the known empirical relationships between a given measure and other variables of interest to the user is the primary determinant of the "goodness" of an instrument. Informative treatments of the process of validation are given in Cronbach (1971) and the Standards for Educational and Psychological Tests (1974).

Table 2 summarizes the major distinctions between the representational point of view advocated by Stevens and the alternative viewpoint suggested in this paper. Following Adams (1966), the alternative is called the "informational" point of view. The informational viewpoint presents a less clearly defined set of assumptions and recommendations than does the representational viewpoint and, consequently, provides fewer concrete guidelines. This is likely to be frustrating to those who seek straightforward solutions to questions about "permissible statistics." Rozeboom (1966) presents a detailed discussion of this

issue in his critique of Stevens but that issue has not been dealt with here.

- - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - -

Realistically, the meaning of a given measure gradually evolves from repeated use and investigation. A measure typically goes through a series of revisions designed to clarify the relevant hypothetical attribute and/or to sharpen the relationships of the measure to other indices which are either predicted from theory or sought in applied decision-making. Several possible procedures which conform to the practical limitations (financial support, time constraints, availability of subjects, technical services available for data analysis, etc.) faced by the instrument developer may be investigated. By successive approximation, a particular operational procedure is defined which serves the intended purpose better than do competing procedures of roughly comparable cost effectiveness. Due to subsequent technological innovations or modifications in theory, a new procedure which is superior to the best existing instrument may be developed. However, it is meaningless to argue the merits of particular psychometric devices in terms of the Stevens' scales. There is no empirical way to resolve such arguments and, even if there were, the question of how well an instrument is serving the purpose for which it was designed would remain unanswered.

References

Adams, E. W. On the nature and purpose of measurement. Synthese, 1966, 16, 125-169.

American Psychological Association. Standards for educational and psychological tests. Washington: APA, 1974.

Campbell, N. R. Physics: the elements, Cambridge, 1920. Reprinted as The Foundations of Science, New York, 1957.

Campbell, N. R. An account of the principles of measurement and calculations. London: Longmans, Green, 1928.

Cronbach, L. J. Test validation. In R. L. Thorndike (Ed.), Educational measurement (2nd Ed.). Washington, D. C.: American Council on Education, 1971, 443-507.

Cronbach, L. J., & Gleser, G. Psychological tests and personal decisions. (2nd Ed.). Urbana: University of Illinois Press, 1965.

Gardner, P. L. Scales and statistics. Review of Educational Research, 1975, 45 (NO. 1), 43-58.

Jones, L. V. The nature of measurement. In R. L. Thorndike (Ed.), Educational measurement (2nd Ed.). Washington, D. C.: American Council on Education, 1971, 335-355.

Nunnally, J. C. Psychometric theory. New York: McGraw-Hill Co., 1967.

Pfanzagl, J. Theory of measurement. New York: Wiley, 1968.

Rozeboom, W. W. Scaling theory and the nature of measurement. Synthese, 1966, 16, 170-233.

Scott, D., & Suppes, P. Foundational aspects of theories of measurement. J. Symb. Logic, 1958, 23.

Shannon, C. E., & Weaver, W. The mathematical theory of communication. Urbana: University of Illinois Press, 1949.

Stevens, S. S. On the theory of scales of measurement. Science, 1946, 103, 667-680.

Stevens, S. S. Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951, 1-49.

Suppes, P., & Zinnes, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology, Vol. 1, New York: Wiley, 1963, 1-76.

Table 1

Distinctions Among Systems
for Indexing Attributes of Objects

	Type of System		
	Classification System	Ranking System	Measurement System
Defining Characteristic	Use of a set of mutually exclusive categories which are unordered.	Use of a set of mutually exclusive and exhaustive categories which are ordered (greater or less)	Use of a well-defined set of operations which result in a numeric index based upon a unit of measurement.
Purpose	To distinguish objects according to some observable (or hypothetical) attribute.	To order objects according to some observable (or hypothetical) attribute.	To index the quantity of some hypothetical attribute possessed by objects.
Examples	- Clinical diagnoses and typologies - College major field (history, math, physics, etc.) - Job classification systems - "Qualitative" individual differences (e.g. sex, blood type, religious affiliation, national origin, etc.)	- Rank in class - Crude indices of socio-economic status - Year in school (fresh, soph, jr, sr) - Job rating or level (senior operator, operator, trainee) - Any sorting or ranking procedure which results in an ordering of objects	- Grade point average - Tests of ability, achievement, aptitude - Attitude, interest, and personality inventories - Likert or Thurstone rating scales - Any "test" which results in a numeric "score" for each individual who responds to it.

Table 2

Some Contrasts in Representational
vs. Informational Points of View

Issue	Representational Viewpoint	Informational Viewpoint
Purpose of Measurement	Measurement procedures should map real world phenomena into number systems such that an isomorphism is established. The more isomorphic that mapping, the more fundamental the level of measurement.	Measurement procedures should provide quantitative indices of real world phenomena that contain useful information. The more evidence that an index is useful in decision-making or in establishing lawful relationships, the better the procedure.
Assumptions Concerning Real World Phenomena	Phenomena can be represented by a variety of operations but some operations are more "fundamental" and hence more desirable. Fundamental operations provide definitions in terms of units and scales that are "true" or "real".	All operations for indexing phenomena are somewhat arbitrary but some provide more useful information than do others. There is no ultimate or "fundamental" operation which defines phenomena in terms of "true" or "real" units and scales.
Systems for Classifying Measurement Procedures	Depending upon how well a procedure reflects real world phenomena, it is classified as nominal, ordinal, interval or ratio. In principle, the level of measurement for a given procedures can be determined by establishing the relationship between that procedure and the "fundamental" operation.	Procedures vary with regard to the potential amount of information they provide. Qualitative classification and rank ordering systems usually have less information potential than do metric procedures and should not be regarded as measurement. However, validation of metric procedures is more complex.
Recommendations Concerning "Permissible Statistics"	Only algebraic operations (transformations)which are invariant with regard to the scale type are permissible. nominal - any substitution of labels that preserves the identity of each class ordinal - any monotonic transformation interval - any linear transformation ratio - any transformation where each scale value is multipled by a constant	It is impossible to specify precise rules for determining whether an algebraic operation is "permissible". For classification and ranking procedures, only operations which yield results that have a "meaningful" interpretation should be employed (see Rozeboom, 1966). For measurement procedures, any algebraic operation should be considered permissible. The use of an operation is justified by subsequent demonstration of lawful relationships or improvement in decision-making